sample 1 sample 2

assume $\rightarrow \begin{bmatrix} \vec{x}_1 & \vec{x}_2 & \vec{x}_3 \end{bmatrix}$

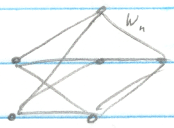
$$z = wx + b$$

3 cols of 2 elements ~~2 cols of 3 elements~~

$$\frac{\partial z^{(1)}}{\partial w^{(1)}} = \begin{bmatrix} \frac{\partial z^{(1)}}{\partial W_1^{(1)}} & \frac{\partial z^{(1)}}{\partial W_2^{(1)}} & \frac{\partial z^{(1)}}{\partial W_3^{(1)}} \end{bmatrix}$$

$$\frac{\partial z^{(1)}}{\partial W_1^{(1)}} = \begin{bmatrix} \frac{\partial z_1^{(1)}}{\partial W_{11}^{(1)}} & \frac{\partial z_1^{(1)}}{\partial W_{12}^{(1)}} \\ & \end{bmatrix}$$

Try looking at it from summation instead of vectors!:




$z_{ij}$ = row i, col j

$$z_{11}^{(1)} = \left(\sum_{i=1}^{2} a_{i1} w_{1i}\right) + b_{11}$$

$$z_{11}^{(1)} = a_{11} w_{11} + a_{21} \cdot w_{12} + b_{11}$$

$$a_{11}^{(1)} = \sigma(z_{11})$$

$$z_{11}^{(2)} = a_{11}^{(2)} \cdot w_{11}^{(2)} + a_{21}^{(2)} \cdot w_{12}^{(2)} + a_{31}^{(2)} \cdot w_{13}^{(2)} + b_{11}^{(2)}$$

general,

$$z_{j1}^{(1)} = \sum_{i=1}^{2} a_{i1}^{(0)} \cdot w_{ji}^{(1)} + b_{j1}^{(1)}$$

$$z_{j1}^{(2)} = \sum_{i=1}^{3} a_{i1}^{(1)} \cdot w_{ji}^{(2)} + b_{j1}^{(2)}$$

$$a_{j1}^{(k)} = \sigma(z_{j1}^{(k)})$$

prediction = $a_{11}^{(2)}$ target = $t_{11}$

Loss = $-t_{11} \log(a_{11}^{(2)}) - (1 - t_{11}) \log(1 - a_{11}^{(2)})$ = Cost

$$\frac{\partial J}{\partial W^{(2)}} = \begin{bmatrix} \frac{\partial J}{\partial W_{11}^{(2)}} & \frac{\partial J}{\partial W_{12}^{(2)}} & \frac{\partial J}{\partial W_{13}^{(2)}} \end{bmatrix} = \begin{bmatrix} a_{11}^{(2)} & a_{21}^{(2)} & a_{31}^{(2)} \end{bmatrix} = a^{(2)T} \cdot \delta$$

$$\frac{\partial J}{\partial W_{11}^{(2)}} = \frac{\partial J}{\partial a_{11}^{(2)}} \cdot \frac{\partial a_{11}^{(2)}}{\partial z_{11}^{(2)}} \cdot \frac{\partial z_{11}^{(2)}}{\partial W_{11}^{(2)}} =$$

$$\frac{\partial J}{\partial a_{11}^{(2)}} = \frac{-t_{11} + a_{11}^{(2)}}{a_{11}^{(2)}(1 - a_{11}^{(2)})}$$

single value

$$\frac{\partial a_{11}^{(2)}}{\partial z_{11}^{(2)}} = \sigma(z_{11}^{(2)})(1 - \sigma(z_{11}^{(2)}))$$

= single value

$$\frac{\partial z_{11}^{(2)}}{\partial W_{11}^{(2)}} = a_{11}^{(2)}$$

$$\frac{\partial z_{11}^{(2)}}{\partial W_{12}^{(2)}} = a_{21}^{(2)}$$

$$\frac{\partial z_{11}^{(2)}}{\partial W_{13}^{(2)}} = a_{31}^{(2)}$$

Now, consider n examples; $z_{11}^{(2)}, z_{12}^{(2)}, \dots, z_{1n}^{(2)}$

$\frac{\partial J}{\partial a_{11}^{(2)}}$ becomes $\begin{bmatrix} \frac{\partial J}{\partial a_{11}^{(2)}} & \cdots & \frac{\partial J}{\partial a_{1n}^{(2)}} \end{bmatrix}$

$\frac{\partial J}{\partial W^{(2)}}$ becomes $\delta_{1 \times n} \cdot a^{(2)T}$

matmul